Toward Conditional Distribution Calibration in Survival Prediction

Shi-ang Qi[1], Yakun Yu[1], Russell Greiner[1,2]

[1]University of Alberta, [2]Alberta Machine Intelligence Institute {shiang, yakun2, rgreiner}@ualberta.ca

Objectives

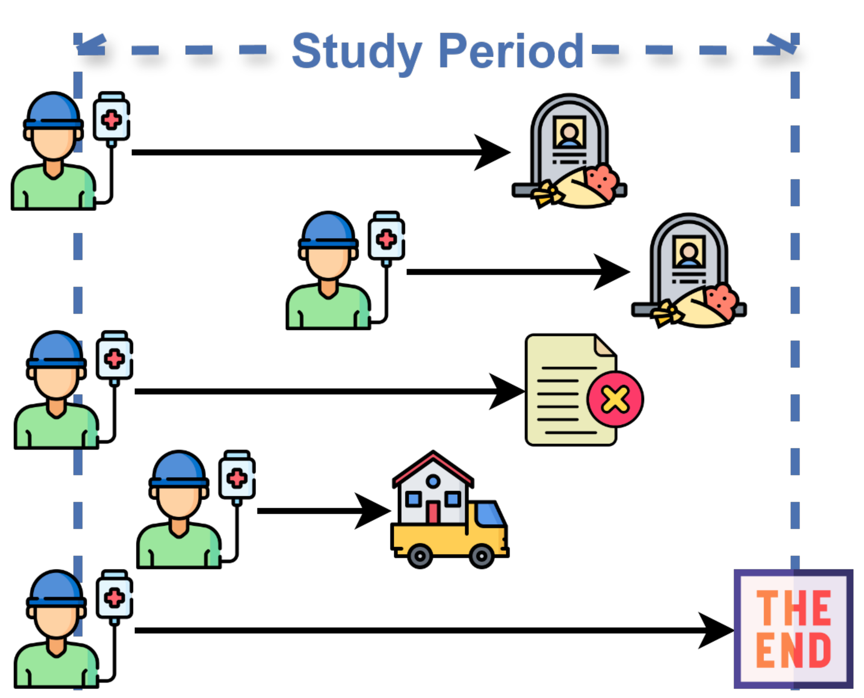
Boost a survival model's marginal and conditional calibration performance while maintaining the same discrimination ability.

Survival Analysis

A subject (described $\boldsymbol{x}_i$) is **right-censored** iff it has not experienced an event at the observed time. Each subject is: $[\boldsymbol{x}_i$, observed time t_i, indicator $\delta_i]$, which is based on event time e_i and censor time c_i.

$$t_i \triangleq \min\{e_i, c_i\} \quad \text{and} \quad \delta_i \triangleq \mathbf{1}[e_i \le c_i]$$

Assumptions: (i) **exchangeable** and (ii) **conditional independent censoring**, $e_i \perp c_i \mid \boldsymbol{x}_i$



Individual Survival Distribution (ISD) is a probability curve for all future times for a patient:

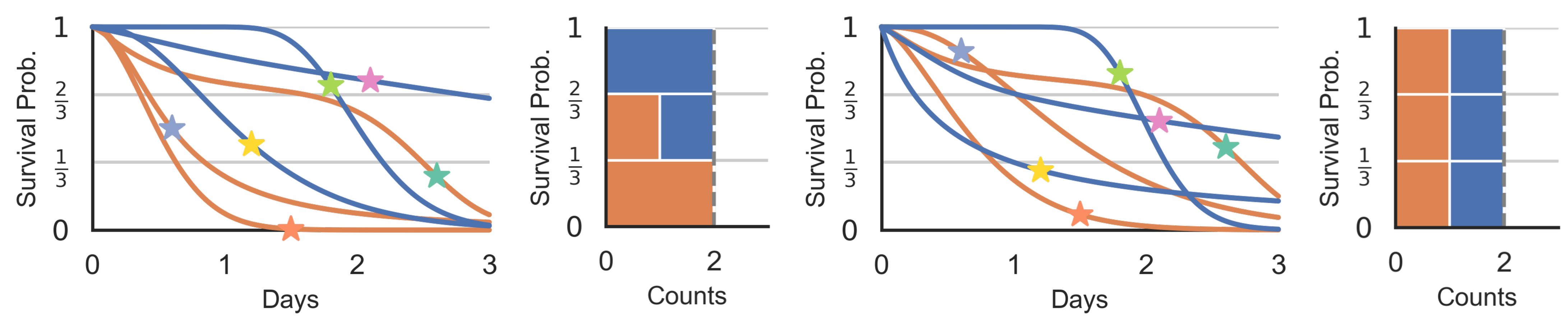
$$S(t \mid \boldsymbol{x}_i) = \Pr(e_i > t \mid \boldsymbol{x}_i).$$

Theoretical Results

Methods	CSD [2]	CSD-iPOT
Marginal calibration guar.†	✗	✓
Conditional calibration guar.	✗	✓
Monotonic	✗	✓
Harrell discrimination guar.	✓	✗
Antolini discrimination guar.	✗	✓
Space complexity‡	$O(N\|\mathcal{P}\|)$	$O(NR)$

† All the calibration guarantees are asymptotic guarantees.
‡ N: #instances in the conformal set; R: sampling parameter; $|\mathcal{P}|$: #predefined percentile.

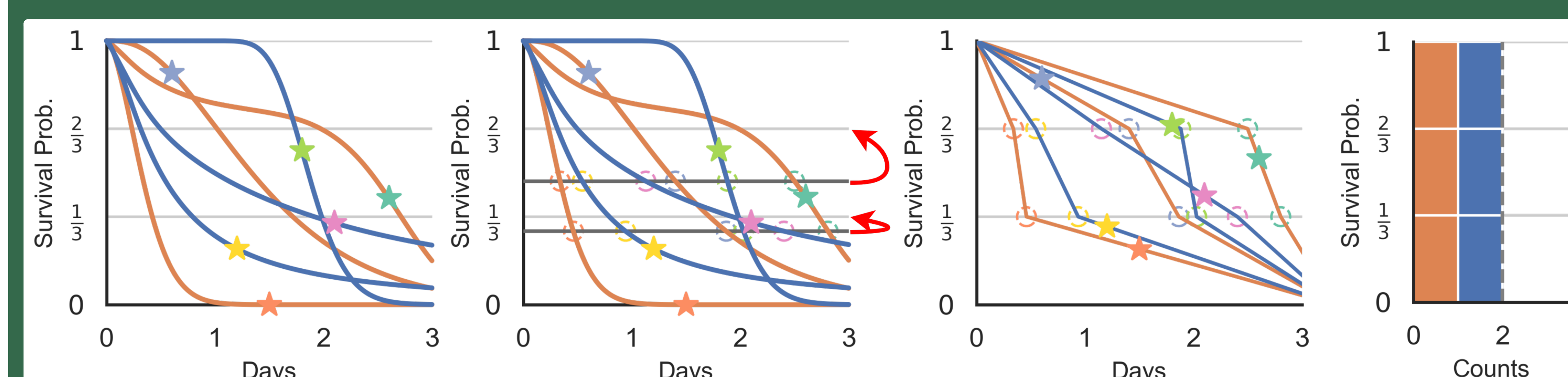
Calibration in Survival Analysis



Marginal calibration[1]: the predicted survival probability at true event time, $\{\hat{S}(e_i \mid \boldsymbol{x}_i)\}_i$, should follow $\mathcal{U}[0,1]$ (*inverse transform theorem*).
For a censored subject, it follows $\mathcal{U}[0, \hat{S}(c_i \mid \boldsymbol{x}_i)]$.

Conditional calibration: $\{\hat{S}(e_i \mid \boldsymbol{x}_i)\}_i$, should follow $\mathcal{U}[0,1]$, for arbitrary group based on $\boldsymbol{x}_i$, e.g., age / sex / race.
We propose Cal_{ws} – evaluate cal. on the worst calibrated sub-region in the feature space.

Method



CSD-iPOT (conformalized survival distribution using individual probability at observed time):

❶ Split data, learn $\mathcal{M}$ from $\mathcal{D}_{\text{train}}$ and predict ISDs for $\mathcal{D}_{\text{con}}$: $\{\hat{S}_{\mathcal{M}}(t \mid \boldsymbol{x}_i)\}_{i\in\mathcal{I}^{\text{con}}}$ (**curves**)

❷ Calculate individual probability at observed time (iPOT) as the conformity score (**stars**)

$$\gamma_{i,\mathcal{M}} := \hat{S}_{\mathcal{M}}(e_i \mid \boldsymbol{x}_i), \quad \Gamma_{\mathcal{M}} = \{\gamma_{i,\mathcal{M}}\}_{i\in\mathcal{I}^{\text{con}}}$$

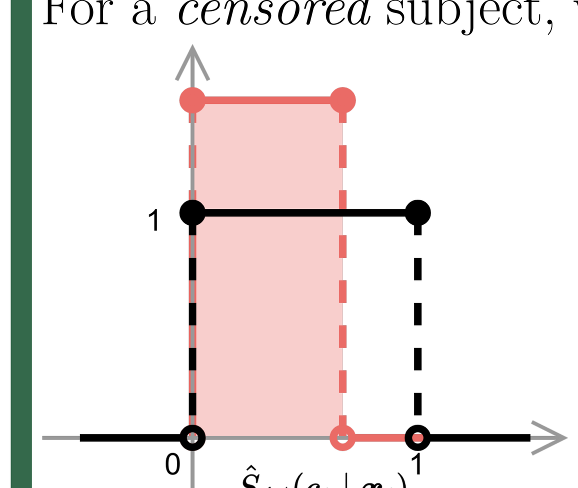
❸ Apply the following adjustment for a testing subject with index $n+1$,

$$\forall\, \rho \in \mathcal{P}, \quad \tilde{S}^{-1}_{\mathcal{M}}(\rho \mid \boldsymbol{x}_{n+1}) = \hat{S}^{-1}_{\mathcal{M}}\left(\mathrm{Percentile}(\rho;\, \Gamma_{\mathcal{M}}) \mid \boldsymbol{x}_{n+1}\right)$$

- ❶ Identify the empirical percentiles of the conformity score (**lines**)
- ❷ Determines the corresponding times on the predicted ISDs that match these empirical percentiles (**circles**)
- ❸ Vertically shift the empirical percentiles to the appropriate height (**arrows**)

❹ Transform the inverse ISD into an ISD: $\tilde{S}_{\mathcal{M}}(t \mid \boldsymbol{x}_{n+1}) = \inf\{\rho : \tilde{S}^{-1}_{\mathcal{M}}(\rho \mid \boldsymbol{x}_{n+1}) \le t\}$

For a *censored* subject, we **cannot** directly calculate conformity score $\gamma_{i,\mathcal{M}} = \hat{S}_{\mathcal{M}}(e_i \mid \boldsymbol{x}_i)$.



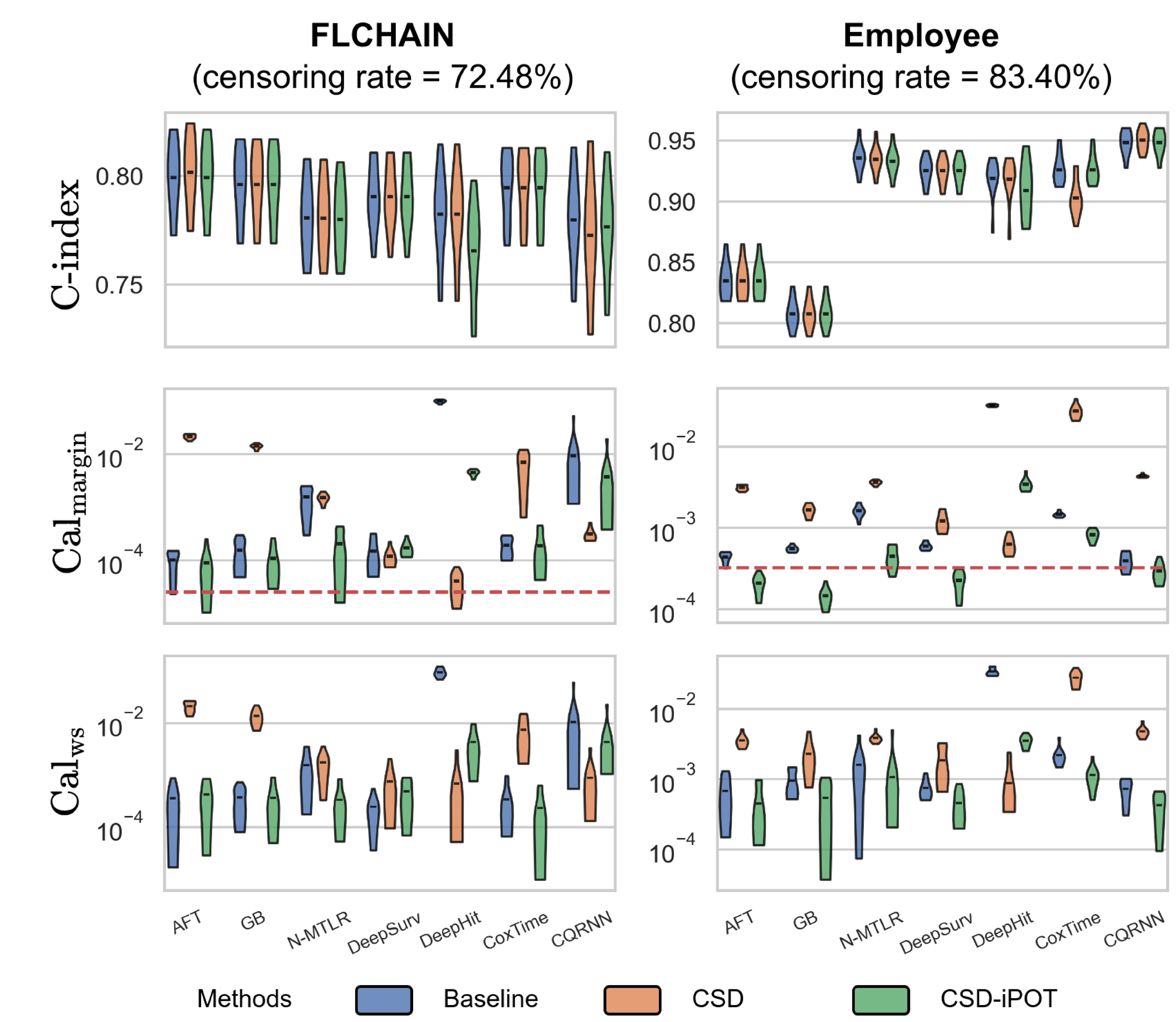
Intuition: given the prior knowledge $\hat{S}_{\mathcal{M}}(e_i \mid \boldsymbol{x}_i) \sim \mathcal{U}_{[0,1]}$, we update the knowledge by $\hat{S}_{\mathcal{M}}(c_i \mid \boldsymbol{x}_i) > \hat{S}_{\mathcal{M}}(e_i \mid \boldsymbol{x}_i)$ and assumption (ii).

Solution: draw R *potential conformity scores* $\sim \mathcal{U}_{[0, \hat{S}_{\mathcal{M}}(c_i|\boldsymbol{x}_i)]}$

$$\gamma_{i,\mathcal{M}} = \hat{S}_{\mathcal{M}}(c_i \mid \boldsymbol{x}_i) \cdot \boldsymbol{u}_R, \quad \text{where} \quad \boldsymbol{u}_R = [0/R, 1/R, \ldots, R/R]$$

To maintain a balanced censoring rate, we repeat the iPOT value, R times, for each uncensored subject.

Empirical Results



Comparisons using 15 real datasets and 7 baselines.

		C-index	$\mathrm{Cal}_{\mathrm{margin}}$	$\mathrm{Cal}_{\mathrm{ws}}$ ‡	IBS	MAE-PO
cf. Baseline	Win	7(0)†	**95(50)**	**64(29)**	**63(14)**	**54(8)**
	Lose	22(0)	9(1)	5(1)	23(0)	17(0)
	Tie	**75**	0	0	18	33
cf. CSD [2]	Win	11(1)	**68(37)**	**51(26)**	**53(15)**	**39(8)**
	Lose	26(0)	36(20)	18(7)	35(11)	39(4)
	Tie	**67**	0	0	16	26

† Number of wins (Number of significant wins with $p < 0.05$).
‡ We only evaluate $\mathrm{Cal}_{\mathrm{ws}}$ on datasets with $n \ge 1000$..

Other findings from ablation studies:

- CSD-iPOT requires less space and running time.
- A larger sampling parameter R can lead to better marginal and conditional calibration.
- Different values of ρ have minimal impacts.

References

[1] Haider et al. Effective ways to build and evaluate individual survival distributions. JMLR 2020
[2] Qi et al. Conformalized Survival Distributions: A Generic Post-Process to Increase Calibration. ICML 2024